Universität St.Gallen

What is the value of knowing the propensity score for estimating average treatment effects?

Markus Frölich

April 2002 Discussion paper no. 2002-06

Department of Economics University of St. Gallen

Editor:
Prof. Jörg Baumberger
University of St. Gallen
Department of Economics
Bodanstr. 1
CH-9000 St. Gallen
Phone ++41 71 224 22 41
Fax ++41 71 224 28 85
Email joerg.baumberger@unisg.ch

Publisher:
Forschungsgemeinschaft für Nationalökonomie
an der Universität St. Gallen
Dufourstrasse 48
CH-9000 St. Gallen
Phone ++41 71 224 23 00
Fax ++41 71 224 26 46

Electronic Publication:
www.fgn.unisg.ch/public/public.htm

What is the value of knowing the propensity score for estimating average treatment effects?

Markus Frölich[1]

Author's address: Markus Frölich
Swiss Institute for International Economics and Applied Economic Research (SIAW)
Dufourstrasse 48
CH-9000 St. Gallen
Tel. ++41 71 2242342
Fax ++41 71 2242298
Email markus.froelich@unisg.ch
Website www.siaw.unisg.ch/froelich, www.markusfroelich.de

[1] I am grateful to Michael Lechner for valuable discussions, comments and suggestions. This research was supported by the Swiss National Science Foundation, project NSF 4043-058311.

Abstract

Propensity score matching is widely used in treatment evaluation to estimate average treatment effects. Nevertheless, the role of the propensity score is still controversial. Since the propensity score is usually unknown and has to be estimated, the efficiency loss arising from not knowing the true propensity score is examined. Hahn (1998) derived the asymptotic variance bounds for known and unknown propensity scores. Whereas the variance of the average treatment effect is unaffected by knowledge of the propensity score, the bound for the treatment effect on the treated changes if the propensity score is known. However, the reasons for this remain unclear. In this paper it is shown that knowledge of the propensity score does not lead to a 'dimension reduction'. Instead, it enables a more efficient estimation of the distribution of the confounding variables.

Keywords

Evaluation, matching, causal effect, semiparametric efficiency bound

JEL Classification

C13, C14

1 Introduction

Propensity score matching is a technique widely used in biometrics, econometrics and other social sciences to estimate the effects of treatment receipt.[1] Its popularity stems from the fact that, instead of controlling for all confounding factors X, it suffices to control for a *one-dimensional* propensity score to remove all selection bias (Rosenbaum and Rubin 1983). Exploiting this 'dimension-reduction' property, many studies used matching on the propensity score to estimate average treatment effects.

However, in applications the propensity score (which is the conditional probability of treatment receipt) is usually unknown, and therefore propensity score matching proceeds on an estimated propensity score. This gave rise to a debate on whether matching on the propensity score is preferable to matching on all confounding variables X (Heckman, Ichimura, and Todd 1998). Substantial research has recently been devoted to clarify the role of the propensity score, and a particular question is, how much less precise is matching on the estimated propensity score, instead of matching on the true propensity score? In other words, what is the value of knowing the true propensity score?

Hahn (1998) derived the $\sqrt{n}$-asymptotic variance bounds for nonparametric estimation of the average treatment effect and the average treatment effect on the treated, as well for known as for unknown propensity score. He found that the variance bound for estimating the average treatment effect (ATE) is the same with and without knowledge of the true propensity score. On the other hand, knowledge of the propensity score reduces the asymptotic variance for the average treatment effect on the treated (ATET). He showed that nonparametric imputation estimators can attain these bounds. Hirano, Imbens, and Ridder (2000) demonstrated that Horvitz and Thompson (1952)-type estimators of average treatment effects can attain these bounds, too. Heckman, Ichimura, and Todd (1998) analyzed local polynomial matching estimators of the treatment effect on the treated and derived their asymptotic distributions, which have a variance term corresponding to the variance bound of Hahn (1998).[2]

[1] See for example, Brodaty, Crépon, and Fougère (2001), Dehejia and Wahba (1999), Frölich, Heshmati, and Lechner (2000), Gerfin and Lechner (2000), Heckman, Ichimura, and Todd (1997), Heckman, Ichimura, Smith, and Todd (1998), Imbens (2000), Jalan and Ravallion (2002), Larsson (2000), Lechner (1999), Puhani (1999), Rosenbaum and Rubin (1983, 1985) and Rubin and Thomas (1992, 1996) among others.

[2] Abadie and Imbens (2001) analyze the asymptotic efficiency of k-nearest-neighbours matching estimators, when k does not grow with increasing sample size (as it is the case with the common pair-matching estimator).

However, the reason *why* knowledge of the propensity score affects the variance bound of the ATET but not of the ATE has not been fully understood. Hahn (1998) argues that the reduction in the variance of the ATET "can be solely attributed to the 'dimension reduction' feature of the propensity score". Yet, this reasoning fails to explain, why the dimension reduction through knowledge of the propensity score has no effect on the variance of the ATE. If indeed the lower dimension of the propensity score would help to avoid the curse of dimensionality, knowing the propensity score should be advantageous for estimating the ATET as well as for estimating the ATE.

In this paper, I provide a different explanation to clarify why knowledge of the propensity score affects the variance bound of the ATET but not of the ATE. I argue that the propensity score is ancillary for the estimation of the conditional expected treatment outcomes $E[Y|X]$, but that it is informative for estimating the *distribution function of the confounding variables X in the treated subpopulation.* This distribution function is the weighting function for the *average* treatment effect on the treated (ATET). If the propensity score is unknown, the distribution function $F_{X|treated}$ is identified by the observations X of the treated individuals only. The control observations are not useful for estimating the distribution $F_{X|treated}$. On the other hand, if the propensity score is known, the control observations are informative to estimate the distribution $F_{X|treated}$, since the propensity score is proportional to the density ratio of X among treated and control. Hence in addition to the treated observations, also the control observations can be used to estimate $F_{X|treated}$. For example in the case of random assignment (with probability 0.5), the distribution of X is the same among the treated and the control, and thus the number of effective observations which can be used to estimate $F_{X|treated}$ increases from $n_{treated}$ to $2n_{treated}$ when it is known that the individuals were randomly assigned.

From this it is apparent, why knowledge of the propensity score has no effect on the variance bound of the ATE, because it is the effect for the whole population and not for a subpopulation. The estimation of the average treatment effect is based on weighting the conditional potential outcomes by the distribution F_X of X in the whole population. Since F_X is naturally estimated from *all* control and *all* treated observations, no other subpopulations can be linked up via the

They demonstrate that these estimators do not attain the asymptotic variance bound. Ichimura and Linton (2001) derived higher order expansions for the Horvitz and Thompson (1952)-type estimator that can be useful for choosing the bandwidth parameter.

propensity score to estimate F_X.

To support this explanation, I show that the different variance bounds of the ATET (with and without knowledge of the propensity score) have the same structure as the variance bounds for estimating the distribution function $F_{X|treated}$. Hence knowledge of the propensity score affects the precision the distribution function $F_{X|treated}$ can be estimated with. Finally it is shown, that if the propensity score is known, the counterfactual outcome *for the treated* can be estimated nonparametrically from the control observations, even if not a single treated observation is available. This is not possible, if the propensity score is unknown.

Hence the *only* value of knowing the propensity score is that it helps to estimate the distribution of the confounding variables X among the treated more precisely. From an asymptotic perspective, matching on the propensity score does not lead to any dimension reduction.

2 Efficiency bounds and the propensity score

Treatment evaluation aims to estimate the causal effects of a particular treatment (e.g. drug treatment, active labour market programmes, school education) by comparing the situation with and without receipt of the treatment. Define Y_i^0, Y_i^1 as the *potential outcomes* of individual i, where Y_i^0 is the outcome that individual i would realize if not receiving the treatment, and Y_i^1 is the outcome that individual i would realize if receiving the treatment, see Neyman (1923) and Rubin (1974, 1977).[3] For each individual only one of the two potential outcomes Y_i^0, Y_i^1 can be observed, but never both. The average causal impact of the treatment can be measured by the average treatment effect (ATE)

$$\alpha = E[Y^1 - Y^0], \tag{1}$$

which is the difference between the outcome expected when receiving the treatment and the outcome expected when not receiving the treatment, for an individual randomly drawn from

[3] The extension to the case where the treatment consists of different varieties, e.g. different drug variants or different labour market programmes, is straightforward and considered in Imbens (2000) and Lechner (2001). Implicit in the potential outcomes Y_i^0, Y_i^1 notation is the assumption of 'no interference between different units' (Cox 1958, p.19) or stable-unit-treatment-value assumption (SUTVA) as called by Rubin (1980): It is assumed that the potential outcomes Y_i^0, Y_i^1 of individual i are not affected by the treatment receipt of other individuals.

the population. Similarly, the average treatment effect on the treated (ATET)

$$\alpha_T = E[Y^1 - Y^0 | D = 1] \tag{2}$$

is the expected outcome difference for an individual randomly drawn from the subpopulation of treatment recipients. $D_i \in \{0,1\}$ indicates whether an individual received treatment or not. Neither of these two average effects (α or α_T) is identified by observational data, since Y^1 can only be observed for the treatment recipients ($D_i = 1$), whereas Y^0 is only observable for non-recipients ($D_i = 0$). Generally, $E[Y^1|D=0] \neq E[Y^1|D=1]$ and $E[Y^0|D=1] \neq E[Y^0|D=0]$ if individuals select (or are assigned) to treatment in a non-random way. This is the well-known selection problem, which prohibits estimating treatment effects by simple mean comparisons (Rubin 1974, Heckman and Robb 1985, Manski 1993). One approach to avoid selection bias is to compare the observed outcomes conditional on all confounding variables, where the confounding variables are all variables X, that influence treatment assignment *as well as* the potential outcomes.[4] Hence conditional on X, the probability of receiving treatment is stochastically independent ($\perp\!\!\!\perp$) of the potential outcomes:

$$Y^0, Y^1 \perp\!\!\!\perp D\,|X, \tag{3}$$

which is known as *selection on observables* (Barnow, Cain, and Goldberger 1981), *ignorable treatment assignment* (Rosenbaum and Rubin 1983) or *conditional independence assumption* (Lechner 1999). Suppose further that X has the same support in the treated and the untreated (control) subpopulation. This is equivalent to assuming that the probability of treatment receipt is bounded away from 0 and 1:

$$0 < \; p(x) \equiv P\,(D=1|X=x) \; < 1 \qquad \forall x \in Supp(X), \tag{4}$$

where $p(x)$ is referred to as the *propensity score* (Rosenbaum and Rubin 1983), i.e. the propensity to receive treatment given characteristics X.

Assumptions (3) and (4) identify the treatment effects α and α_T: Since the expected potential outcomes conditional on X are independent of treatment status by (3), α and α_T are identified by an iterated expectations argument.[5] Estimation of $E[Y^0]$ and $E[Y^0|D=1]$

[4] In a randomized experiment with full compliance the set of confounding variables is empty.

[5] Because $E[Y^0] = E[E[Y^0|X]] = E[E[Y^0|X,D=1]]$ and $E[Y^0|D=1] = E[E[Y^0|X,D=1]|D=1] = E[E[Y^0|X,D=0]|D=1]$.

thus requires only a proper weighting of the conditional expectation functions by the relevant distribution function of X. Define the conditional mean function

$$m_1(x) \equiv E[Y^1|X = x] = E[Y^1|X = x, D = 1],$$

and $m_0(x)$ analogously. The average treatment effect α and the average treatment effect on the treated α_T can be written as

$$\begin{aligned} E\left[Y^1 - Y^0\right] &= \int (m_1(x) - m_0(x)) \cdot dF_X(x) \qquad (5) \\ E\left[Y^1 - Y^0|D = 1\right] &= \int (m_1(x) - m_0(x)) \cdot dF_{X|D=1}(x), \end{aligned}$$

where $f_X \equiv dF_X$ is the density of X in the whole population, and $f_{X|D=1} \equiv dF_{X|D=1}$ is the density of X in the treated subpopulation. The conditional expectation functions $m_1(x)$ and $m_0(x)$ are separately[6] identified from the treated and the control observations, respectively. The distributions $F_X(x)$ and $F_{X|D=1}(x)$ are identified from the X observations of treated and controls. From (5) it can be seen that the *only difference* between the average treatment effect α and the average treatment effect on the treated α_T is, that in the former the conditional means difference $m_1(x) - m_0(x)$ is weighted by the density dF_X, whereas it is weighted by $dF_{X|D=1}$ in the latter effect. While dF_X is the density of X in the *whole* population, $dF_{X|D=1}$ is the density in a particular *sub*population. As will be seen below, this is the reason why knowledge of the propensity score affects the variance bound of α_T but not of α: If the propensity score is known, $dF_{X|D=1}$ is identified from the treated *and* the control observations, whereas the control observations are not informative for estimating $dF_{X|D=1}$ if the propensity score is unknown. On the other hand, dF_X is always identified from the treated and the control observations together, regardless of knowledge about the propensity score. In other words, knowledge of the propensity score improves upon the estimation of $dF_{X|D=1}$, but does not affect any other term in (5).

Nonparametric estimation of $m_1(x)$ and $m_0(x)$ can be difficult in finite samples if the dimension of X is high, which may often be the case when numerous factors determine treatment receipt (e.g. occurrence of non-compliance in drug trials, self-selection to active labour market programmes). Therefore Rosenbaum and Rubin (1983) introduced the *propensity score* $p(x)$ as

[6] The joint distribution of Y^0, Y^1 is not identified, since for each individual only one of the potential outcomes can be observed.

a way to reduce the dimension of the estimation problem. They showed that if independence of the potential outcomes Y^0, Y^1 and treatment assignment D holds conditional on X, then it does also hold conditional on the probability of treatment receipt $p(x) \equiv P(D=1|X=x)$:

$$Y^0, Y^1 \perp\!\!\!\perp D\,|X \qquad\qquad \Rightarrow \qquad\qquad Y^0, Y^1 \perp\!\!\!\perp D\,|p(X). \tag{6}$$

As a result, the treatment effects α and α_T is also identified as

$$\begin{aligned} E\left[Y^1 - Y^0\right] &= \int (\mathfrak{m}_1(\rho) - \mathfrak{m}_0(\rho)) \cdot dF_p(\rho) \qquad (7) \\ E\left[Y^1 - Y^0|D=1\right] &= \int (\mathfrak{m}_1(\rho) - \mathfrak{m}_0(\rho)) \cdot dF_{p|D=1}(\rho), \end{aligned}$$

where $\mathfrak{m}_1(\rho) \equiv E[Y^1|p(X) = \rho] = E[Y^1|p(X) = \rho, D = 1]$ by (6), and $\mathfrak{m}_0(\rho)$ defined analogously. dF_p is the density distribution of $p(X)$ in the whole population, and $dF_{p|D=1}$ the density distribution in the treated subpopulation. The conditional expectations $\mathfrak{m}_1(\rho)$ and $\mathfrak{m}_0(\rho)$ can be estimated by nonparametric regression on the *one-dimensional* propensity score $p(x)$, and weighting by the corresponding density function of the propensity score yields hence a consistent estimate of the treatment effects α and α_T. In this sense, propensity score matching circumvents the dimensionality problem of direct matching on X, as in (5), and is therefore widely used in applied evaluation studies. In the case where the propensity score is unknown, propensity matching proceeds on an estimated propensity score.

For estimating average treatment effects often a nonparametric imputation estimator is employed. Let $\{(X_i, D_i, Y_i)\}_{i=1}^n$ be an iid sample of size $n = n_0 + n_1$, where n_0 is the number of control observations and n_1 is the number of treated observations. Consider the average treatment effect on the treated α_T. A method of moment estimator of α_T based on (5) is

$$\frac{1}{n_1} \sum (Y_i - \hat{m}_0(X_i)) \cdot D_i,$$

where $\hat{m}_0$ is a nonparametric estimate of m_0, obtained from the n_0 control observations. The analogous propensity score matching estimator is

$$\frac{1}{n_1} \sum (Y_i - \hat{\mathfrak{m}}_0(\hat{p}(X_i))) \cdot D_i,$$

where $\hat{\mathfrak{m}}_0$ is an estimate of $\mathfrak{m}_0$ and $\hat{p}(x)$ is the (estimated) propensity score. Often $\hat{m}_0$ and $\hat{\mathfrak{m}}_0$ are estimated by first-nearest-neighbour regression (=pair-matching).[7] However, matching

[7] This is the reason, why these estimators are called 'matching' estimators, since first-nearest-neighbour regression $\hat{m}_0(X_i)$ matches pairs of treated and control observations.

estimators based on first-nearest-neighbour regression are inefficient (Abadie and Imbens 2001), and Heckman, Ichimura, and Todd (1998) analyzed local polynomial regression estimators $\hat{m}_0$ as a more efficient alternative.[8]

To analyze the efficiency of treatment effect estimators, Hahn (1998) derived the *semiparametric efficiency bounds* for nonparametric estimation of α and α_T, as well for known as for unknown propensity score.[9] The variance bound of the average treatment effect α is

$$\frac{1}{n_0+n_1}\cdot E\left[\frac{\sigma_1^2(X)}{p(X)}+\frac{\sigma_0^2(X)}{1-p(X)}+(m_1(X)-m_0(X)-\alpha)^2\right], \tag{8}$$

where $\sigma_1^2(x) = Var\left(Y^1|X=x\right) = Var\left(Y^1|X=x, D=1\right)$. The multiplicative term $\frac{1}{n_0+n_1}$ in front of (8) is attached to demonstrate that the variance of the estimated treatment effect vanishes at rate $n^{-1}=(n_0+n_1)^{-1}$.

This variance bound of α is the same for known and for unknown propensity score. In other words, knowledge of the true propensity is not informative for estimating α. Furthermore, a projection on the propensity score (i.e. matching on the propensity score as in (7)) is not necessary to attain the variance bound (8), as shown in Hahn (1998). Thus from an asymptotic perspective, propensity score matching (7) and matching on X as in (5) are equivalent, even if the propensity score is known.

On the other hand, the variance bound of the average treatment effect on the treated α_T depends on the knowledge of the propensity score. If the propensity score is *unknown*, the variance bound of α_T is

$$\frac{1}{n_0+n_1}\frac{1}{P^2}\cdot E\left[\sigma_1^2(X)p(X)+\frac{\sigma_0^2(X)p(X)^2}{1-p(X)}+p(X)\left(m_1(X)-m_0(X)-\alpha_T\right)^2\right] \tag{9}$$

where $P=P\left(D=1\right)=\lim\frac{n_1}{n_0+n_1}$ is the fraction of treated individuals. When the propensity score is *known*, the variance bound of α_T is

$$\frac{1}{n_0+n_1}\frac{1}{P^2}\cdot E\left[\sigma_1^2(X)p(X)+\frac{\sigma_0^2(X)p(X)^2}{1-p(X)}+p^2(X)\left(m_1(X)-m_0(X)-\alpha_T\right)^2\right]. \tag{10}$$

Again, a projection on the propensity score is not necessary to attain these bounds, i.e. asymptotically propensity score matching (7) does not provide any advantage to matching on X as in

[8] Frölich (2000) compared the finite-sample properties of various propensity score matching estimators and found that significant precision gains are possible vis-a-vis first-nearest-neighbour regression.

[9] Semiparametric efficiency bounds were introduced by Stein (1956) and developed by Koshevnik and Levit (1976), Pfanzagl and Wefelmeyer (1982), Begun, Hall, Huang, and Wellner (1983) and Bickel, Klaassen, Ritov, and Wellner (1993). See also the survey of Newey (1990).

(5). Nevertheless, Hahn attributes the reduction of the variance from (9) to (10) to the 'dimension reducing' property of the propensity score. As a heuristic argument, he notes that knowledge of the propensity score implies knowledge of the equivalence classes $\mathcal{X}_c = \{x : p(x) = c\}$, which are the sets of x with the same propensity score value. As an extreme example for the dimension reduction due to knowledge of the equivalence classes, he considers the case of random treatment assignment, which implies $\alpha = \alpha_T$, $p(x) = p$ for all individuals and $\mathcal{X}_p = Supp(X)$. In this case the bounds (8) and (10) are identical, whereas the variance bound of α_T with unknown propensity score (9) is larger, see Theorem 3 in Hahn (1998). The difference between (9) and (10) is attributed by Hahn to the dimension reduction property of the propensity score.

However, this interpretation is rather misleading. First of all, if the propensity score would indeed contribute to a dimension reduction, it should also improve the estimation of α. A more suitable explanation seems to be, that knowledge of the true propensity score helps in estimating the distribution function $F_{X|D=1}$ of X in the treated subpopulation. Without knowledge of the propensity score, the distribution $F_{X|D=1}$ can only be estimated from the X observations of the n_1 treated individuals. The X values of the control observations contain no information about $F_{X|D=1}$. On the other hand, if the propensity score is known, also the X observations of the control individuals are informative about the distribution of X among the treated, since the propensity score relates the density of X between the treated and the control subpopulations: By Bayes' theorem $p(x) = f_{X|D=1}(x)P(D = 1)/f_X(x)$ and thus

$$\frac{p(x)}{1 - p(x)} = \frac{f_{X|D=1}(x)}{f_{X|D=0}(x)} \frac{P(D = 1)}{P(D = 0)}, \tag{11}$$

where $f_{X|D=0}$ is the density in the control subpopulation. With this relationship all $n_0 + n_1$ observations can be used to estimate $F_{X|D=1}$.

This also explains why knowledge of the propensity score is of no use for estimating the average treatment effect α, which is the average outcome difference $m_1 - m_0$ weighted by the distribution F_X of X in the population, see (5). Since F_X can be estimated from the empirical distribution function of all $n_0 + n_1$ observations, its estimation cannot be improved upon by knowledge of the propensity score. Hence the value of the propensity score is, that it enables using the X observations of one subpopulation (the control) to estimate the distribution of X in a different subpopulation (the treated). Consequently, the impact of knowing the true propensity score on the variance bound is zero, when the latter 'subpopulation' is the whole population (as in the estimation of the population average treatment effect α.)

By this explanation it is also obvious why the variance bounds (8) and (10) coincide in the random assignment setting ($p(x) = p$), and why (9) does not (as noticed in Theorem 3 of Hahn (1998) and discussed above). With random assignment $\alpha = \alpha_T$ and the distribution of X is identical among the treated and the control: $F_{X|D=1} = F_{X|D=0} = F_X$. Estimation of α would proceed by estimating F_X from the empirical distribution function of all $n_0 + n_1$ observations and computing (5). The estimation of α_T depends on the knowledge about the propensity score. If the propensity score is known, the distribution $F_{X|D=1}$ can be estimated from the empirical distribution function of all $n_0 + n_1$ observations, too, since (11) implies $f_{X|D=1}(x)/f_{X|D=0}(x) = 1$. Hence the bounds (8) and (10) coincide. On the other hand, if the propensity score is unknown, the distribution $F_{X|D=1}$ is estimated from the empirical distribution function of the n_1 treated observations, and the X observations of the controls are completely neglected, although they have the same distribution of X. The estimator of α_T with unknown propensity score neglects n_0 of the available observations for estimating $F_{X|D=1}$. Thus the variance of α_T vanishes at rate $\frac{1}{n_1}$ instead of $\frac{1}{n_0 - n_1}$, which explains the difference between (9) and (10).[10] Knowledge of the propensity does not lead to a dimension reduction. It simply allows a more efficient estimation of $F_{X|D=1}$.

To examine the case of non-random assignment ($p(x) \neq p$), the three variance bounds (8) to (10) can be rewritten as

$$\frac{1}{n_1} E_{f_1}\left[\sigma_1^2(X) \frac{f_X^2(X)}{f_{X|D=1}^2(X)}\right] + \frac{1}{n_0} E_{f_0}\left[\sigma_0^2(X) \frac{f_X^2(X)}{f_{X|D=0}^2(X)}\right] + \frac{1}{n_0 + n_1} E\left[(m_1(X) - m_0(X) - \alpha)^2\right] \tag{12a}$$

$$\frac{1}{n_1} E_{f_1}\left[\sigma_1^2(X)\right] + \frac{1}{n_0} E_{f_0}\left[\sigma_0^2(X) \frac{f_{X|D=1}^2(X)}{f_{X|D=0}^2(X)}\right] + \frac{1}{n_1} E_{f_1}\left[(m_1(X) - m_0(X) - \alpha_T)^2\right] \tag{12b}$$

$$\frac{1}{n_1} E_{f_1}\left[\sigma_1^2(X)\right] + \frac{1}{n_0} E_{f_0}\left[\sigma_0^2(X) \frac{f_{X|D=1}^2(X)}{f_{X|D=0}^2(X)}\right] + \frac{1}{n_0 + n_1} E_{f_1}\left[\frac{f_{X|D=1}(X)}{f_X(X)} (m_1(X) - m_0(X) - \alpha_T)^2\right] \tag{12c}$$

where $E_{f_1}[\cdot] = \int \cdot f_{X|D=1}(x)dx$ refers to the expected value in the treated subpopulation, $E_{f_0}[\cdot]$ to the expected value in the control subpopulation, and $P = P(D = 1) = \lim \frac{n_1}{n_0 + n_1}$ is approximated by $\frac{n_1}{n_0 + n_1}$. The first row corresponds to the bound (8) of the average treatment effect

[10] Noting that $\frac{1}{n_1} - \frac{1}{n_0+n_1} = \frac{1}{n_0+n_1}\frac{n_0}{n_1} \simeq \frac{1}{n_0+n_1}\frac{1-P}{P}$, which corresponds to the difference between (9) and (10): $\frac{1}{n_0+n_1}\frac{1-P}{P} \cdot E\left[(m_1(X) - m_0(X) - \alpha_T)^2\right]$ when $p(x) = P$.

α (with and without knowledge of the propensity score). The second row refers to the average treatment effect on the treated α_T with *unknown* propensity score (9), and the third row refers to the bound of α_T with *known* propensity score (10). The first term in each of these bounds captures the variance due to estimating $m_1(x)$, re-weighted by the density of X in the relevant population. This term vanishes at rate $\frac{1}{n_1}$, since only the n_1 treated observations are informative for estimating m_1. Analogously, the second term represents the variance due to estimating $m_0(x)$ and vanishes at rate $\frac{1}{n_0}$, since only the control observations can be used for estimating m_0. The third term stems from estimating the distributions F_X and $F_{X|D=1}$, respectively, and vanishes either at rate $\frac{1}{n_1}$ or $\frac{1}{n_0+n_1}$.

Since knowledge of the propensity score does not affect the first two terms of the bounds (12b) and (12c), the only channel through which the propensity score influences the variance bound is through the third term, which corresponds to the estimation of $F_{X|D=1}$. To ease the following discussion, consider the case where the conditional expectation functions m_0 and m_1 are known. Then the first two terms of (12a)-(12c) are zero,[11] and the corresponding variance bound for α is

$$\frac{1}{n_0+n_1} \cdot E\left[\left(m_1(X) - m_0(X) - \alpha \right)^2 \right], \tag{13}$$

and the bounds for α_T (without and with knowledge of the the propensity score) are

$$\frac{1}{n_1} \cdot \underset{f_1}{E} \left[\left(m_1(X) - m_0(X) - \alpha_T \right)^2 \right] \tag{14a}$$

$$\frac{1}{n_0+n_1} \cdot \underset{f_1}{E} \left[\frac{f_{X|D=1}(X)}{f_X(X)} \left(m_1(X) - m_0(X) - \alpha_T \right)^2 \right]. \tag{14b}$$

The variance of α vanishes at rate $\frac{1}{n_0+n_1}$, since the distribution of X in the population is identified from all observations. On the other hand, the main difference between the variance bound of α_T with knowledge (14b) and without knowledge of the propensity score (14a) is, that the variance vanishes at rate $\frac{1}{n_1}$ in (14a), while it vanishes at rate $\frac{1}{n_0+n_1}$ in (14b). Again the reason for this is, that with unknown propensity score the distribution $F_{X|D=1}$ is identified from treated observations only, whereas treated *and* control observations are informative for $F_{X|D=1}$ when the propensity score is known, such that its estimation can be based on all n_1+n_0 observations. If the distribution of X differs in the control and the treated subpopulation (non-random assignment), the control observations are less 'efficient' in estimating the distribution

[11] This follows immediately by repeating the proofs of Hahn (1998).

$F_{X|D=1}$,[12] which is embodied in the correction term $f_{X|D=1}/f_X$ in (14b). In case of random assignment ($F_{X|D=1} = F_X$), the control observations are as 'efficient' in estimating $F_{X|D=1}$ as the treated observations, and the bounds (14b) and (13) agree. The variance bound of the average treatment effect α is unaffected by the propensity score, since it is based on the distribution F_X in the *whole* population, such that no other subpopulations can be linked up for its estimation via the propensity score.

This relationship between the propensity score and the distribution function $F_{X|D=1}$ becomes even more apparent, when examining the variance bounds of the estimated distribution function $F_{X|D=1}(a)$. When the propensity score is *unknown*, the variance bound for estimating $F_{X|D=1}(a)$ is

$$\frac{1}{n_0+n_1}\cdot E\left[\frac{p(x)}{P^2}\left[1\left(x\le a\right)-F_{X|D=1}(a)\right]^2\right]$$
$$=\frac{1}{n_1}\cdot \underset{f_1}{E}\left[\left(1\left(X\le a\right)-F_{X|D=1}(a)\right)^2\right], \quad (15)$$

and when the propensity score is *known* it is

$$\frac{1}{n_0+n_1}\cdot E\left[\frac{p^2(x)\left[1\left(x\le a\right)-F_{X|D=1}(a)\right]^2}{P^2}\right]$$
$$=\frac{1}{n_0+n_1}\cdot \underset{f_1}{E}\left[\frac{f_{X|D=1}(X)}{f_X(X)}\left[1\left(X\le a\right)-F_{X|D=1}(a)\right]^2\right], \quad (16)$$

(Proof in Appendix). These variance bounds have the same structure as the bounds (14a) and (14b). If the propensity score is unknown, the variance (15) vanishes at rate $\frac{1}{n_1}$, whereas the variance (16) vanishes at rate $\frac{1}{n_0+n_1}$ for known propensity score, because in the latter case the control observations assist in estimating $F_{X|D=1}$ (again with $f_{X|D=1}/f_X$ as correction factor as in (14b)). Hence knowledge of the propensity score makes a more efficient estimator of the distribution $F_{X|D=1}(a)$ available.

This estimator uses the empirical distribution function of the control and the treated observations, to estimate the distribution $F_{X|D=1}$. Consider the estimation of the counterfactual mean outcome $E[Y^0|D=1]$, which is the crucial ingredient in the average treatment effect on

[12] Because the density mass of the control observations may be located to a large extent in different regions than the mass of the treated observations.

the treated α_T.[13] The common matching estimator of $E[Y^0|D=1] = \int m_0(x) \cdot dF_{X|D=1}(x)$ is

$$E\,[\widehat{Y^0|D}=1] = \frac{1}{n_1} \sum_{i:D_i=1} \hat{m}_0(X_i), \tag{17}$$

where $\hat{m}_0(x)$ is estimated from the control observations and imputed for all treated observations $(D_i = 1)$. However, if the propensity score is known, the efficient estimator of $E[Y^0|D=1]$ is

$$E\,[\widehat{Y^0|D}=1] = \frac{\frac{1}{n_1+n_0}\sum\limits_i \hat{m}_0(X_i)p(X_i)}{\frac{1}{n_1+n_0}\sum\limits_i p(X_i)}, \tag{18}$$

which, in contrast to (17), is a weighted average of m_0 for the treated *and* the control observations. The estimator (18) is motivated on

$$E[Y^0|D=1] = \int m_0(x) f_{X|D=1}(x)dx = \int m_0(x)\frac{p(x)f_X(x)}{P(D=1)}dx = \frac{\int m_0(x)p(x)f_X(x)dx}{\int p(x)f_X(x)dx} \tag{19}$$

and corresponds to the efficient estimator in Proposition 7 of Hahn (1998).

The value of knowing the propensity score for estimating $F_{X|D=1}$ becomes even more obvious, when rewriting (19) by using (11)[14] as

$$E[Y^0|D=1] = \int m_0(x)\frac{p(x)}{1-p(x)}\frac{P(D=0)}{P(D=1)} f_{X|D=0}(x)dx = \frac{\int m_0(x)\frac{p(x)}{1-p(x)} f_{X|D=0}(x)dx}{\int \frac{p(x)}{1-p(x)} f_{X|D=0}(x)dx},$$

for $p(x) \neq 1$, which suggests the estimator

$$E\,[\widehat{Y^0|D}=1] = \frac{\frac{1}{n_0}\sum\limits_{i:D_i=0} \hat{m}_0(X_i)\frac{p(X_i)}{1-p(X_i)}}{\frac{1}{n_0}\sum\limits_{i:D_i=0} \frac{p(X_i)}{1-p(X_i)}}. \tag{20}$$

Although the estimator (20) is inefficient, it demonstrates that, with knowledge of the propensity score, the counterfactual mean for the treated $(E[Y^0|D=1])$ can be estimated from the control observations, even *without* a single treated observation available, because the control observations identify the distribution function $F_{X|D=1}$. This is not possible, if the propensity score is unknown.

Hence from an asymptotic point of view, knowing the propensity score leads to a more efficient estimator of the distribution function of X among the treated, but it does not contribute to any dimension reduction.

[13] Because $E[Y^1|D=1]$ can be estimated simply by the sample mean outcome of the treated.

[14] And that $\int \frac{p(x)}{1-p(x)} f_{X|D=0}(x)dx = \int p(x)\frac{f_X(x)}{f_{X|D=0}(x)P(D=0)} f_{X|D=0}(x)dx = \frac{1}{P(D=0)}\int p(x)f_X(x)dx = \frac{P(D=1)}{P(D=0)}$.

A Appendix

Below the semiparametric variance bound for the estimation of the distribution of X among the treated, $F_{X|D=1}(a)$, is derived, following closely the proof in Hahn (1998). Examine first the case where the propensity score is unknown.

The joint density of (D, X) can be written as

$$f(d,x) = f(d|x)f(x) = p(x)^d(1-p(x))^{1-d}f(x). \tag{21}$$

Consider a regular parametric submodel indexed by θ

$$f(d,x,\theta) = p(x,\theta)^d(1-p(x,\theta))^{1-d}f(x,\theta),$$

such that $f(d,x,\theta_0) = f(d,x)$. The score of $f(d,x,\theta)$ is

$$\begin{aligned} S(d,x,\theta) &= \frac{\partial \ln f(d,x,\theta)}{\partial\theta} = d\frac{\partial p(x,\theta)/\partial\theta}{p(x,\theta)} - (1-d)\frac{\partial p(x,\theta)/\partial\theta}{1-p(x,\theta)} + \frac{\partial f(x,\theta)/\partial\theta}{f(x,\theta)} \\ &= \frac{d-p(x,\theta)}{p(x,\theta)\,(1-p(x,\theta))}\frac{\partial p(x,\theta)}{\partial\theta} + \frac{\partial f(x,\theta)/\partial\theta}{f(x,\theta)} \\ &= \frac{d-p(x,\theta)}{p(x,\theta)\,(1-p(x,\theta))}\dot{p}(x,\theta) + \frac{\dot{f}(x,\theta)}{f(x,\theta)}, \end{aligned}$$

where $\dot{p}(x,\theta) = \frac{\partial p(x,\theta)}{\partial\theta}$ and $\dot{f}(x,\theta) = \frac{\partial f(x,\theta)}{\partial\theta}$

The tangent space of the model is

$$\Im = \{\varphi(x)\cdot(d-p(x,\theta)) + s(x)\}$$

for all square-integrable functions $\varphi(x)$ and all functions $s(x)$ satisfying $\int s(x)f(x)dx = 0$.

The semiparametric variance bound of $F_{X|D=1}(a)$ is the expected square of the projection on $\Im$ of a function $\mathfrak{D}(d,x)$, which satisfies for all regular parametric submodels

$$\frac{\partial F_{X|D=1}(a,\theta)}{\partial\theta}_{|\theta=\theta_0} = E\left[\mathfrak{D}(D,X)\cdot S(D,X,\theta)|\theta=\theta_0\right]. \tag{22}$$

The distribution function $F_{X|D=1}(a,\theta)$ can be written as

$$F_{X|D=1}(a,\theta) = E_\theta\left[1\left(X\le a\right)|D=1\right] = \int 1\left(x\le a\right)f_{X|D=1}(x,\theta)dx = \frac{\int 1\left(x\le a\right)p(x,\theta)f(x,\theta)dx}{\int p(x,\theta)f(x,\theta)dx}$$

since $p(x) = f_{X|D=1}(x)P/f(x)$ with $P = E[D] = \int p(x)f(x)dx$. Its pathwise derivative is

$$\frac{\partial F_{X|D=1}(a,\theta)}{\partial\theta}_{|\theta=\theta_0} = \frac{\left(\int 1\left(x \le a\right)\left(\frac{\partial p(x,\theta)}{\partial\theta}f(x,\theta) + \frac{\partial f(x,\theta)}{\partial\theta}p(x,\theta)\right)dx\right)P}{P^2} \tag{23}$$

$$- \frac{\left(\int 1\left(x \le a\right)p(x,\theta)f(x,\theta)dx\right)\left(\int\left(\frac{\partial p(x,\theta)}{\partial\theta}f(x,\theta) + \frac{\partial f(x,\theta)}{\partial\theta}p(x,\theta)\right)dx\right)}{P^2}_{|\theta=\theta_0} \tag{24}$$

$$= \frac{\int 1\left(x \le a\right)\left(\frac{\partial p(x,\theta_0)}{\partial\theta}f(x) + \frac{\partial f(x,\theta_0)}{\partial\theta}p(x)\right)dx}{P} \tag{25}$$

$$- \frac{F_{X|D=1}(a)\int\left(\frac{\partial p(x,\theta_0)}{\partial\theta}f(x) + \frac{\partial f(x,\theta_0)}{\partial\theta}p(x)\right)dx}{P} \tag{26}$$

$$= \frac{\int\left(1\left(x \le a\right) - F_{X|D=1}(a)\right)\left(\frac{\partial p(x,\theta_0)}{\partial\theta}f(x) + \frac{\partial f(x,\theta_0)}{\partial\theta}p(x)\right)dx}{P} \tag{27}$$

$$= \frac{1}{P}\int\left[1\left(x \le a\right) - F_{X|D=1}(a)\right]\left(\dot{p}(x)f(x) + p(x)\dot{f}(x)\right)dx, \tag{28}$$

where $\dot{p}(x) = \frac{\partial p(x,\theta)}{\partial\theta}_{|\theta=\theta_0}$ and $\dot{f}(x) = \frac{\partial f(x,\theta)}{\partial\theta}_{|\theta=\theta_0}$.

Let

$$\mathfrak{D}(d,x) = \frac{d}{P}\left[1\left(x \le a\right) - F_{X|D=1}(a)\right],$$

which satisfies (22) because

$$\begin{aligned}
&E\left[\mathfrak{D}(D,X)\cdot S(D,X,\theta)|\theta=\theta_0\right]\\
&= \sum_{d=0}^{1}\int\left(\frac{d}{P}\left[1\left(x \le a\right) - F_{X|D=1}(a)\right]\right)\left(\frac{d-p(x)}{p(x)\left(1-p(x)\right)}\dot{p}(x) + \frac{\dot{f}(x)}{f(x)}\right)f\left(d,x\right)dx\\
&= \sum_{d=0}^{1}\int\frac{d}{P}\left[1\left(x \le a\right) - F_{X|D=1}(a)\right]\left(\frac{d-p(x)}{p(x)\left(1-p(x)\right)}\dot{p}(x) + \frac{\dot{f}(x)}{f(x)}\right)p(x)^d(1-p(x))^{1-d}f(x)dx\\
&= \frac{1}{P}\int\left[1\left(x \le a\right) - F_{X|D=1}(a)\right]\left(\frac{1-p(x)}{p(x)\left(1-p(x)\right)}\dot{p}(x) + \frac{\dot{f}(x)}{f(x)}\right)p(x)f(x)dx\\
&= \frac{1}{P}\int\left[1\left(x \le a\right) - F_{X|D=1}(a)\right]\left(\dot{p}(x)f(x) + p(x)\dot{f}(x)\right)dx = \frac{\partial F_{X|D=1}(a,\theta)}{\partial\theta}_{|\theta=\theta_0}
\end{aligned}$$

after inserting (21), integrating out d and comparing with (28).

The variance bound is the expected square of the projection of $\mathfrak{D}(d,x)$ on the tangent space $\Im$. Notice that $\mathfrak{D}(d,x) \in \Im$, because it can be written as $\mathfrak{D}(d,x) = \left[1\left(x \le a\right) - F_{X|D=1}(a)\right]\frac{d-p(x,\theta_0)}{P} + \left[1\left(x \le a\right) - F_{X|D=1}(a)\right]\frac{p(x,\theta_0)}{P}$. Hence the projection of

$\mathfrak{D}(d,x)$ is $\mathfrak{D}(d,x)$ and the variance bound (with unknown propensity score) is thus

$$\begin{aligned} E\left[\mathfrak{D}(D,X)^2\right] &= E\left[\frac{D^2}{P^2}\left[1\left(x\leq a\right)-F_{X|D=1}(a)\right]^2\right] \\ &= E\left[\frac{p(x)}{P^2}\left[1\left(x\leq a\right)-F_{X|D=1}(a)\right]^2\right]. \end{aligned}$$

Now, consider the case where the propensity score is known. The parametric submodel changes to

$$f\left(d,x,\theta\right)=p(x)^d(1-p(x))^{1-d}f(x,\theta),$$

and the score is

$$S(d,x,\theta)=\frac{\partial\ln f\left(d,x,\theta\right)}{\partial\theta}=\frac{\partial f(x,\theta)/\partial\theta}{f(x,\theta)}.$$

The tangent space of the model is thus

$$\Im=\left\{s(x):\int s(x)f(x)dx=0\right\}.$$

Compute the pathwise derivative of

$$F_{X|D=1}(a,\theta)=\frac{\int 1\left(x\leq a\right)p(x)f(x,\theta)dx}{\int p(x)f(x,\theta)dx}$$

$$\begin{aligned} \frac{\partial F_{X|D=1}(a,\theta)}{\partial\theta}_{|\theta=\theta_0} &= \frac{\left(\int 1\left(x\leq a\right)p(x)\frac{\partial f(x,\theta)}{\partial\theta}dx\right)P}{P^2}-\frac{\left(\int 1\left(x\leq a\right)p(x)f(x,\theta)dx\right)\left(\int p(x)\frac{\partial f(x,\theta)}{\partial\theta}dx\right)}{P^2}_{|\theta=\theta_0} \\ &= \frac{\int 1\left(x\leq a\right)p(x)\frac{\partial f(x,\theta_0)}{\partial\theta}dx}{P}-\frac{F_{X|D=1}(a)\left(\int p(x)\frac{\partial f(x,\theta_0)}{\partial\theta}dx\right)}{P} \\ &= \frac{1}{P}\int\left[1\left(x\leq a\right)-F_{X|D=1}(a)\right]p(x)\left(\frac{\partial f(x,\theta_0)/\partial\theta}{f(x,\theta_0)}\right)f(x,\theta_0)dx \\ &= E\left[\left(\frac{p(X)\left[1\left(X\leq a\right)-F_{X|D=1}(a)\right]}{P}\right)S(D,X,\theta_0)|\theta=\theta_0\right]. \end{aligned}$$

Hence

$$\mathfrak{D}(d,x)=\frac{p(x)\left[1\left(x\leq a\right)-F_{X|D=1}(a)\right]}{P}$$

with $\mathfrak{D}(d,x)\in\Im$. Accordingly the variance bound of $F_{X|D=1}(a)$ with known propensity score is

$$E\left[\mathfrak{D}(D,X)^2\right]=E\left[\frac{p^2(x)\left[1\left(x\leq a\right)-F_{X|D=1}(a)\right]^2}{P^2}\right].$$

References

ABADIE, A., AND G. IMBENS (2001): "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects," mimeo, Harvard University.

BARNOW, B., G. CAIN, AND A. GOLDBERGER (1981): "Selection on Observables," *Evaluation Studies Review Annual*, 5, 43–59.

BEGUN, J., W. HALL, W. HUANG, AND J. WELLNER (1983): "Information and Asymptotic Efficiency in Parametric-Nonparametric Models," *Annals of Statistics*, 11, 432–452.

BICKEL, P., C. KLAASSEN, Y. RITOV, AND J. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. John Hopkins University Press, Baltimore.

BRODATY, T., B. CRÉPON, AND D. FOUGÈRE (2001): "Using matching estimators to evaluate alternative youth employment programmes: Evidence from France, 1986-1988," in *Econometric Evaluation of Labour Market Policies*, ed. by M. Lechner, and F. Pfeiffer, pp. 85–124. Physica/Springer, Heidelberg.

COX, D. (1958): *Planning of Experiments*. Wiley, New York.

DEHEJIA, R., AND S. WAHBA (1999): "Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programmes," *Journal of American Statistical Association*, 94, 1053–1062.

FRÖLICH, M. (2000): "Nonparametric Covariate Adjustment: Pair-matching versus Local Polynomial Matching," *University of St. Gallen Economics Discussion Paper Series*, 2000-17.

FRÖLICH, M., A. HESHMATI, AND M. LECHNER (2000): "A Microeconometric Evaluation of Rehabilitation of Long-term Sickness in Sweden," *University of St. Gallen Economics Discussion Paper Series*, 2000-04.

GERFIN, M., AND M. LECHNER (2000): "Microeconometric Evaluation of the Active Labour Market Policy in Switzerland," *University of St. Gallen Economics Discussion Paper Series*, 2000-10.

HAHN, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331.

HECKMAN, J., H. ICHIMURA, J. SMITH, AND P. TODD (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017–1098.

HECKMAN, J., H. ICHIMURA, AND P. TODD (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 64, 605–654.

——— (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294.

Heckman, J., and R. Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labour Market Data*, ed. by J. Heckman, and B. Singer. Cambridge University Press, Cambridge.

Hirano, K., G. Imbens, and G. Ridder (2000): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *NBER, Technical Working Paper*, 251.

Horvitz, D., and D. Thompson (1952): "A Generalization of Sampling without Replacement from a Finite Population," *Journal of American Statistical Association*, 47, 663–685.

Ichimura, H., and O. Linton (2001): "Asymptotic Expansions for Some Semiparametric Program Evaluation Estimators," mimeo, University College London.

Imbens, G. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions," *Biometrika*, 87, 706–710.

Jalan, J., and M. Ravallion (2002): "Estimating the Benefit Incidence of an Antipoverty Program by Propensity Score Matching," *Journal of Business and Economic Statistics*, fortcoming.

Koshevnik, Y., and B. Levit (1976): "On a Non-parametric Analogue of the Information Matrix," *Theory of Probability and Applications*, 21, 738–753.

Larsson, L. (2000): "Evaluation of Swedish Youth Labour Market Programmes," *Scandinavian Working Papers in Economics*, 2000:1.

Lechner, M. (1999): "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification," *Journal of Business and Economic Statistics*, 17, 74–90.

——— (2001): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption," in *Econometric Evaluation of Labour Market Policies*, ed. by M. Lechner, and F. Pfeiffer, pp. 43–58. Physica/Springer, Heidelberg.

Manski, C. (1993): "The Selection Problem in Econometrics and Statistics," in *Handbook of Statistics*, ed. by G. Maddala, C. Rao, and H. Vinod. Elsevier Science Publishers.

Newey, W. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135.

Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles.," *Statistical Science*, Reprint, 5, 463–480.

Pfanzagl, J., and W. Wefelmeyer (1982): *Contributions to a General Asymptotic Statistical Theory*. Springer Verlag, Berlin.

PUHANI, P. (1999): *Evaluating Active Labour Market Policies: Empirical Evidence for Poland during Transition*. Physica, Heidelberg.

ROSENBAUM, P., AND D. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

——— (1985): "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *The American Statistician*, 39, 33–38.

RUBIN, D. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

——— (1977): "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2, 1–26.

——— (1980): "Comment on 'Randomization Analysis of Experimental Data: The Fisher Randomization Test' by D. Basu," *Journal of American Statistical Association*, 75, 591–593.

RUBIN, D., AND N. THOMAS (1992): "Characterizing the Effect of Matching using Linear Propensity Score Methods with Normal Distributions," *Biometrika*, 79, 797–809.

——— (1996): "Matching using Estimated Propensity Scores: Relating Theory to Practice," *Biometrics*, 52, 249–264.

STEIN, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, Berkeley.